

# OpenKG Chain: A Blockchain Infrastructure for Open Knowledge Graphs

Huajun Chen<sup>1†</sup>, Ning Hu<sup>2</sup>, Guilin Qi<sup>3</sup>, Haofen Wang<sup>4</sup>, Zhen Bi<sup>1</sup>, Jie Li<sup>2</sup> & Fan Yang<sup>1</sup>

<sup>1</sup>College of Computer Science & AZFT Knowledge Engine Lab, Zhejiang University, Hangzhou 310058, China

<sup>2</sup>Ontology Foundation, Venture Drive #11-31, Vision Exchange, Singapore 608526, Singapore

<sup>3</sup>School of Computer Science and Engineering, Southeast University, Nanjing 211189, China

<sup>4</sup>Intelligent Big Data Visualization Lab, Tongji University, Shanghai 200092, China

**Keywords:** Open knowledge graph; Blockchain; Decentralized distributed network; Decentralized knowledge graph

Citation: Chen, H.J., et al.: OpenKG chain: A blockchain infrastructure for Open Knowledge Graphs. Data Intelligence 3(2), 205-227 (2021). doi: 10.1162/dint\_a\_00095

Received: January 10, 2021; Revised: March 15, 2021; Accepted: April 25, 2021

## ABSTRACT

The early concept of knowledge graph originates from the idea of the semantic Web, which aims at using structured graphs to model the knowledge of the world and record the relationships that exist between things. Currently publishing knowledge bases as open data on the Web has gained significant attention. In China, Chinese Information Processing Society of China (CIPS) launched the OpenKG in 2015 to foster the development of Chinese Open Knowledge Graphs. Unlike existing open knowledge-based programs, OpenKG chain is envisioned as a blockchain-based open knowledge infrastructure. This article introduces the first attempt at the implementation of sharing knowledge graphs on OpenKG chain, a blockchain-based trust network. We have completed the test of the underlying blockchain platform, and the on-chain test of OpenKG's data set and tool set sharing as well as fine-grained knowledge crowdsourcing at the triple level. We have also proposed novel definitions: K-Point and OpenKG Token, which can be considered to be a measurement of knowledge value and user value. 1,033 knowledge contributors have been involved in two months of testing on the blockchain, and the cumulative number of on-chain recordings triggered by real knowledge consumers has reached 550,000 with an average daily peak value of more than 10,000. For the first time, we have tested and realized on-chain sharing of knowledge at entity/triple granularity level. At present, all operations on the data sets and tool sets at OpenKG.CN, as well as the triplets at OpenBase, are recorded on the chain, and corresponding value will also be generated and assigned in a trusted mode. Via this effort, OpenKG chain looks forward to providing a more credible and traceable knowledge-sharing platform for the knowledge graph community.

<sup>†</sup> Corresponding author: Huajun Chen (Email: huajunsir@zju.edu.cn; ORCID: 0000-0001-5496-7442).

## 1. OPEN KNOWLEDGE ECOSYSTEM

### 1.1 Knowledge Graphs as World Models

Knowledge such as facts, information or descriptions, is the awareness and understanding of the world. Knowledge is acquired through experience or education by perceiving, discovering and learning [1, 2]. The early concept of knowledge graph originates from the idea of semantic Web [3, 4] by Tim Berners Lee who is honored as inventor of the World Wide Web. It aims at using the structured graph to model the knowledge of the world and record the relationships that exist between things in the world [5].

Generally speaking, knowledge graphs (KGs) are directed labeled graph (DLG) structures that capture knowledge in the form of triplets (*subject, predicate, object*), expressed as  $(s, p, o)$ , where  $s$  and  $o$  denote entities, and  $p$  establishes the relationship between entities. Due to the convenience of building semantic connections between real-world objects and domain knowledge, many large-scale business KGs have been built in recent years such as Google and Baidu Knowledge Graph, Microsoft Satori, and Product Knowledge Graph from Alibaba and Amazon. These KGs have led to a broad range of important applications, such as question answering [6], language understanding [7, 8], relational data analysis [9, 10], and recommendation systems [11]. Meanwhile, along with early AI research on knowledge representations such as ontologies [12, 13] and followed consolidation of semantic Web standards such as RDF/OWL [14, 15], the most recent advances in deep representation learning and graph neural networks [16, 17] have led to brand new development of knowledge graph technologies.

### 1.2 Open Knowledge Graphs

Along with the burgeoning of the largest open knowledge sharing media, i.e., the World Wide Web, publishing knowledge bases as open data on the Web has gained significant attention since the early days of the Web. Typical examples include the Linked Open Data efforts which were initiated by the semantic Web community and have already collected over 3,000 public linked data sets, the ConeceptNet [18] which originates from a Web-based, crowdsourced Open Mind Common Sense project launched by MIT media lab since 1999, and Wikidata [19], which is a free and editable knowledge base set up by Wikipedia Foundation. In China, Chinese Information Processing Society of China (CIPS) launched the OpenKG in 2015 to foster the development and openness of Chinese knowledge graphs. OpenKG has accumulated over 200 billion triplets in Chinese since its birth and the size is growing fast. Unlike existing open knowledge base programs, OpenKG chain is envisioned as a blockchain-based open knowledge infrastructure, which will be introduced in detail in the following sections.

### 1.3 Value Chain of Open Knowledge

Knowledge is a valuable resource. As illustrated in Figure 1, the production, transformation, exchange, and consumption of knowledge form the value chain of the knowledge in society. Upon the open Web infrastructure, it is a real challenge to build a trust value chain to support the lifecycle of knowledge.

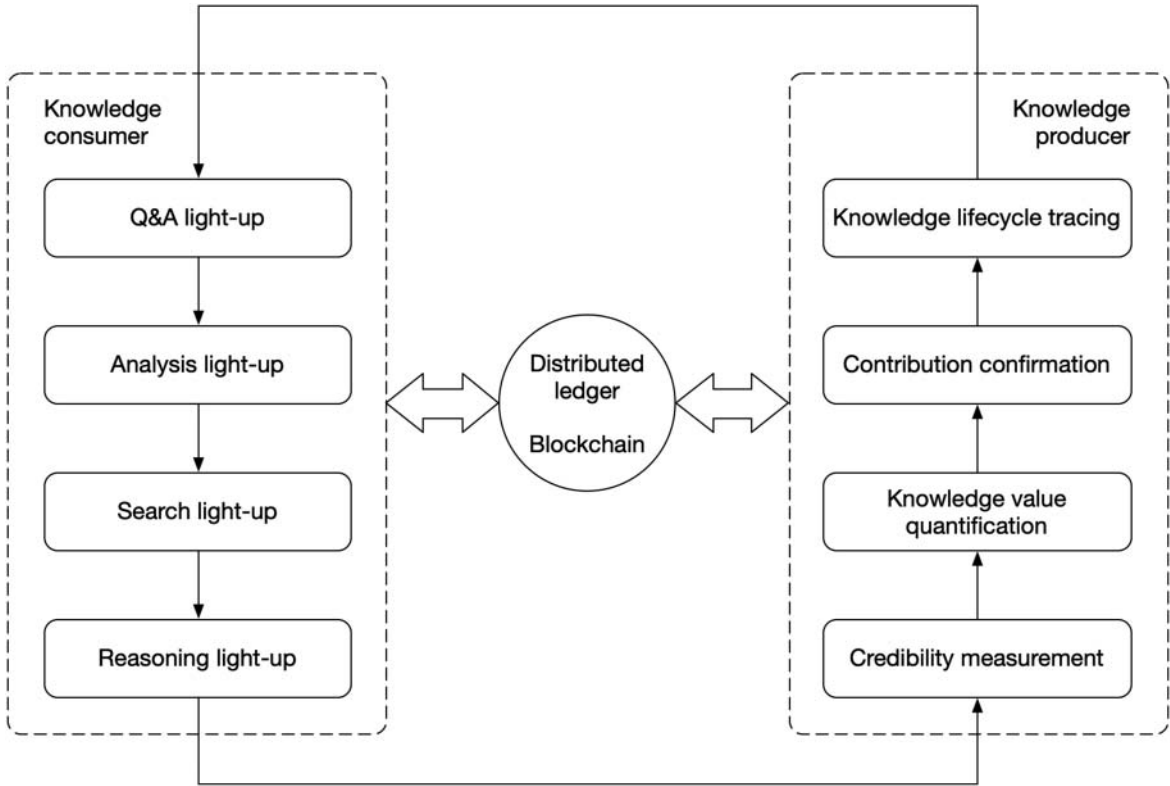


Figure 1. Open ecosystem of knowledge graphs.

### 1.3.1 Incentive Knowledge at a Triple Granularity

The backend logic between contribution and incentives in a society can be described vividly. The more contributions we make, the more incentives we get. It motivates people to contribute more to the society with better quality. According to Maslow's hierarchy of needs, the incentive is not limited to incomes, but better to be measurable. As people share their own knowledge and create social value, it is possible to evaluate and have incentives by knowledge directly. The Web, like a giant social media, cannot satisfy the requirement to track, evaluate, validate the contribution of sharing knowledge and make incentives. Given a knowledge graph, another challenge is to do knowledge-based incentives at a triple granularity. From the delivery of factual knowledge in a triple format, every step of its verification, consumption, transmission, and deletion is traceable and measurable. The basic requirement of a robust sharing platform is to evaluate the value of knowledge in triple form, track its contribution during knowledge processing or consumption, and grant proper incentive in the whole lifecycle of the knowledge.

### 1.3.2 Self-sovereign Knowledge

For now, most knowledge graphs are using centralized systems and data are stored in centralized servers. So the contributors cannot make full control of their knowledge, and the ownership of the knowledge goes to centralized systems. OpenKG chain uses the term self-sovereign knowledge as the concept of individuals or organizations to take full responsibility of their knowledge, make full control of the knowledge, and reveal knowledge with privacy protection or copyright. Individuals and organizations with decentralized identifiers (in any KG system) can be discovered, located, and share their knowledge triples to each other without going through an intermediary, even cross systems. OpenKG chain is trying to provide a self-sovereign interlinking KG system with privacy protection and without an intermediary.

### 1.3.3 Adversarial Attack and Knowledge Accountability

In an open media like the Web, everyone can publish statements or contribute/consume knowledge equally. It brings forth a salient concern on holding all stakeholders accountable for either the statements they make or the actions they take on knowledge. Firstly, statement accountability ensures the authenticity of a statement if everyone is aware that he/she is responsible for what is stated. Secondly, action accountability monitors all activities performed upon knowledge and keeps track of the whole lifecycle of a triple from its birth, edits, consumption, and deletion, preventing illegal actions or adversarial attack on the knowledge base. For example, someone may add a malicious rumor statement into a knowledge base or illegally delete a fact he/she is reluctant to reveal to the public.

### 1.3.4 Immutable Knowledge and Tamper Resistance

Another issue relevant to knowledge accountability is Tamper Resistance. In some cases, the knowledge triples are either sensitive or the integrity of the content is critical. Therefore they must be protected from being fraudulently and intentionally modified, and a tamper resistance network infrastructure is thus required. To ensure that once a statement is committed to the network, both the content and all the follow-on transactions made upon the statement cannot be altered or compromised retrospectively. The integrity of the knowledge content is guaranteed and the transactions cannot be tampered with by any means.

### 1.3.5 Lighting-up and Dissemination of Knowledge Value

The consumption of knowledge is the most direct way to measure the value of knowledge. The more knowledge is consumed, the higher is the value of knowledge. Meanwhile, the consumption of knowledge triggers the dissemination of the value of knowledge. We call the process of knowledge being consumed the lighting-up of knowledge value. The usage scenarios of the knowledge graph support different knowledge users to the nodes in the knowledge graph to trigger knowledge spread.

**“Lighting-up by search”** refers to that the knowledge users consume knowledge during the search process, which triggers the value lighting-up of the searched knowledge item. Knowledge graphs support

semantically related search. And further related searches will continue to trigger new knowledge lighting-up. Each step of lighting-up records the value generated. Since knowledge comes from different producers, the produced value also needs to be awarded to the corresponding knowledge producers on the chain in an accountable way.

**“Lighting-up by question and answering”** is similar to “Lighting-up by search”. A question issued by a user triggers the lighting-up of the knowledge triple touched by the question. Meanwhile the intermediate nodes traversed from the starting node to the answer node in the question and answering retrieval process will also be further lighted up and value-recorded.

**“Lighting-up by inference”** refers to the knowledge lighting-up triggered by the inference process. The knowledge in the knowledge graph is usually incomplete, and the reasoning process of the knowledge graph is completed based on the existing knowledge in the knowledge graph. At the same time, due to many sources of knowledge, the process of lighting-up by inference may also be completed federally, that is, lighting-up by federal inference [20, 21, 22].

**“Lighted up by analysis”** refers to the comprehensive analysis of knowledge from different sources to continuously trigger the lighting process of related knowledge in the knowledge graph. Similarly, due to the diversity of knowledge sources, the analysis process may also be complete in a federated manner. For example, an analysis model may be established through federate learning [23, 24].

## 2. WHEN KNOWLEDGE GRAPH MEETS BLOCKCHAIN

### 2.1 Blockchain and Distributed Ledger

Blockchain [25, 26] uses the distributed ledger [27, 28, 29] technology, which is a kind of ledger database shared, copied and synchronized inside an open P2P network. The data storage and processing is completed by each node inside the P2P network. Therefore, each node can participate in monitoring the legitimacy of the transaction and testify for the transaction results. The blockchain constitutes a multi-centralized network with consensus [30, 31, 32] on a complete transaction log and execution results, which is characterized by “immutable”, “traceability”, and “right confirmation”. Based on these characteristics, the blockchain technology has laid a solid “trust” foundation and created a reliable “cooperation” mechanism. An open knowledge platform provides service for collective maintenance of knowledge, to make full trace of knowledge evolution process, to measure the contribution of knowledge contributors quantitatively, and to meet the requirement of data privacy. The development of knowledge graph takes into account the knowledge quantification, iterative history tracing, and the governance of knowledge points, contributors and knowledge development environment. Therefore, OpenKG blockchain is used to satisfy the requirements.

## 2.2 Open Knowledge and Blockchain

The construction of open-domain knowledge graphs reflects the social attributes of the open community which brings forth a variety of challenges:

- Identify more individual roles and avoid oligopoly of open knowledge. It is required to identify the same entity with different roles to participate in collaborative work and clarify the contribution of different roles to the open knowledge network. Furthermore, open knowledge contributors manage their data autonomously to avoid unauthorized misuse caused by data concentration.
- Support for more decentralized trust management and more controllable qualification of domain experts for open knowledge. The qualification of domain experts in different fields is essential for high-quality knowledge crowdsourcing. The levels of qualification recognition need to be adjusted dynamically, and in turn, more quantitative and fine-grained quantitative evaluation programs need to be implemented.
- The ability to quantify the contributions of massive participants. It is required to track the value of open knowledge contributed by a large number of contributors and adjust the knowledge value model based on feedback from massive participants.

By using the distributed ledger technology, the generation, development, and deduction of open knowledge are recorded, the value and ownership of open knowledge can be fully tracked. For instance, a multi-centralized blockchain network provides trusted infrastructure, tracks the process of open knowledge development, and guarantees data authenticity. A decentralized identity system supports multi-dimensional management of distributed data tokens and massive user tokens. The distributed token solution of the blockchain supports the calculation of knowledge value points, reflecting the value of open knowledge.

In summary, the open knowledge graph structured on a decentralized distributed network is bound to face many issues, including incentives, ownership management, traceability, trust, and privacy. However, the existing centralized knowledge graph management platform does not consider these issues, thus discouraging the sharing and interconnection of knowledge, nor can it guarantee the authenticity and timeliness of knowledge. So we propose a blockchain-based open knowledge graph platform, the functional components of which can be sorted into three levels, including knowledge production, knowledge dissemination, and knowledge consumption, as shown in Figures 2 and 3. The knowledge production layer corresponds to traditional technologies such as knowledge modeling, extraction, fusion, and verification. The knowledge dissemination layer needs to consider the fine-grained knowledge rewarding, self-sovereign knowledge management, knowledge accountability, adversarial attack and taper resistance, and data privacy protection. The knowledge consumption layer includes semantic search and question answering, reasoning, federated learning, and process automation such as robotic process automation (RPA) [33, 34] and other series of applications that need to be built on distributed knowledge sources.

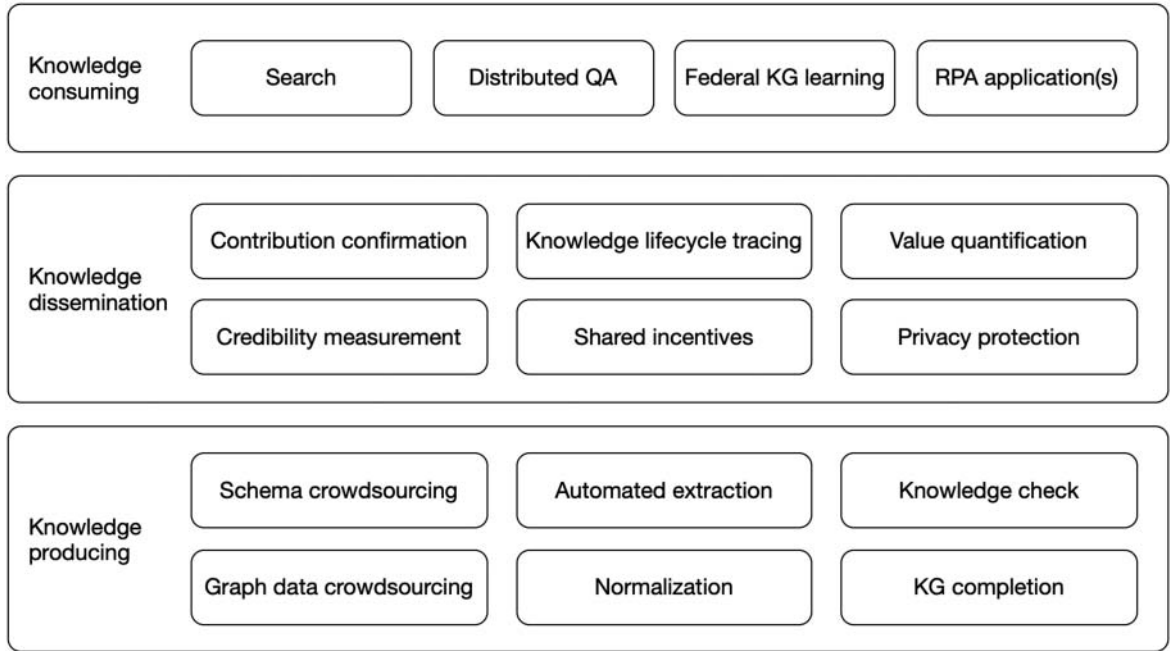


Figure 2. Federated knowledge graph technology platform architecture.

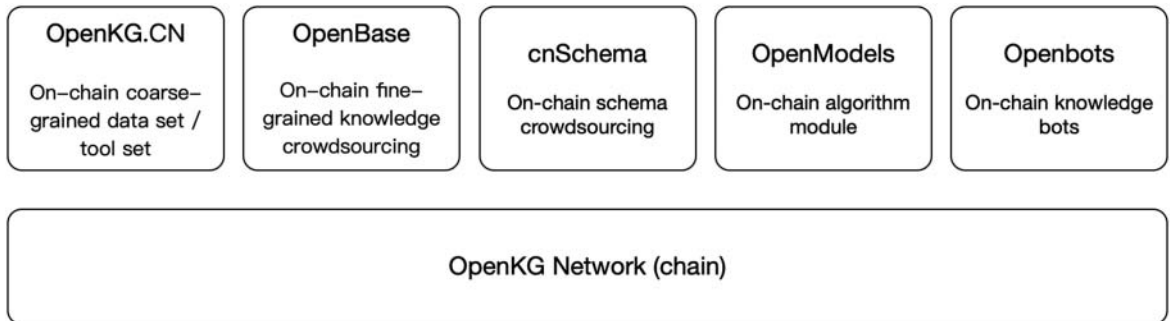


Figure 3. OpenKG resources sharing platform.

### 3. OPENKG BLOCKCHAIN

#### 3.1 OpenKG Resource Model

OpenKG chain consists of several websites aiming at sharing different types of knowledge graph resources.

- The OpenKG main site provides a sharing platform for coarse-grained open resources such as KG data sets and KG tool sets contributed by the KG community in China.
- CnSchema provides a crowdsourced open schema for Chinese knowledge graphs.
- OpenBase is a fine-grained triple-level knowledge graph crowdsourcing platform.



At present, OpenKG chain has completed the construction of the underlying blockchain infrastructure, as well as the on-chain testing of sharing different types of knowledge resources collected by OpenKG community through OpenKG chain. The initial nodes of OpenKG blockchain network are tentatively set to be seven, which are delivered and deployed in different universities or corporate institutions for operation. These seven nodes are independent of each other and form a multi-center blockchain infrastructure for OpenKG community, which builds upon a consensus mechanism to provide a distributed trusted infrastructure. More core nodes can be gradually expanded as needed. In this test platform, there are already more than 1,000 registered knowledge contributors. The two-month average daily value of the on-chain test reaches 10,691 times, and the total number of lights and on-chain deposits exceeds 550,000. It is the first test that has realized the knowledge confirmation of entity/triple granularity (Figure 4).



**Figure 4.** Statistics of the number of times OpenKG lighting-up in May, 2020. Note: Txn Count: KG access count; Unique Senders: KG point dispatchers; Unique Receivers: Unique KG point receivers; Total Unique: The sum of unique senders and unique receivers.

### 3.2 OpenKG Value Model

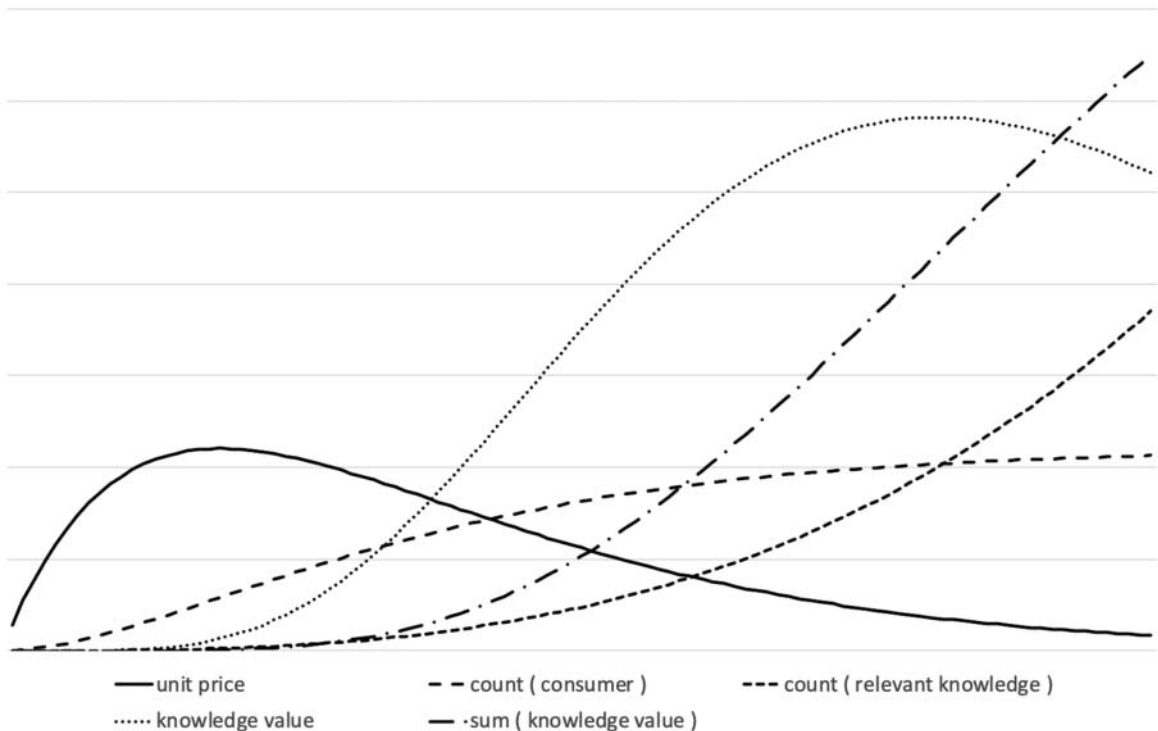
The first issue that the OpenKG chain needs to address is to define proper value models to reflect the value of knowledge. In the case of KG, the value calculation needs to be finely controlled at a triple level. The K-Point is proposed to measure knowledge value for triple knowledge published in OpenKG. Secondly, since OpenKG chain gathers knowledge in the form of community crowdsourcing [35, 36, 37], we also need to design a value model to measure and honor the contribution of knowledge contributors (Table 1).



**Table 1.** Statistics of the number of times OpenKG chain lighting-up in May 2020.

Curve	Characteristics
Knowledge value (unit price)	In the process of the value development of a knowledge unit, when a few people understand it, the unit value is higher. With more and more acceptance and use, the unit value gradually decreases.
Knowledge consumer	Knowledge is limited by the domain, and the number of people who understand gradually increases, and the domain is gradually saturated. The more knowledge audiences, the more knowledge use.
Relevant knowledge points	As knowledge is accepted, it will reason or discover the relationship with other knowledge and form new knowledge. The more relevant knowledge points, the more knowledge will be used.
Value of knowledge	The number of uses of knowledge and the unit price of knowledge form the value of knowledge.
Cumulative value of knowledge	Because of the consistency of knowledge, knowledge has cumulative value.

According to Maslow's hierarchy of needs, the incentive to contribute is not limited to incomes, but better to be measurable. A knowledge value point is proposed to measure the value from a knowledge perspective, OpenKG Token is proposed to measure the contributions and honor the contributors. The hypothesis of the knowledge value model from the very beginning is described as the following, and the knowledge value model can be demonstrated in Figure 5.

**Figure 5.** Knowledge value model.

### 3.2.1 K-Point: Knowledge Value Measurement

OpenKG chain has designed the K-Point contract to reflect the value of knowledge. The assessment of knowledge value is based on a simple model, e.g., each time the knowledge is used, the K-Point is increased accordingly. In current settings, a simple chi-square distribution is used for fitting the model as illustrated below. As knowledge usage scenarios increase, OpenKG chain will continue to use some learnable algorithms to calibrate and optimize the value evaluation models.

$$f(x, v) = \text{Gamma.dist}(x, v / 2, 2) = \frac{1}{2^{v/2} \cdot \Gamma(v / 2)} x^{v/2-1} e^{-x/2} \quad (1)$$

Without considering the interrelationship of knowledge applications, let:

$$\begin{cases} f_{\text{Knowledge value expectation}}(K, x) = f(x, 2) \\ f_{\text{Knowledge use expectation}}(K, x) = f(2x, 5) \end{cases} \quad (2)$$

K is a single knowledge point:

$$f_{\text{Knowledge value quota}}(K, x) = \frac{f_{\text{Knowledge value expectations}}(K, x) \cdot f_{\text{Knowledge use expectation}}(K, x)}{\sum (f_{\text{Knowledge value expectation}}(K, x) \cdot f_{\text{Knowledge use expectation}}(K, x))} \quad (3)$$

Let  $x \in (0, 10]$ ,  $\text{Count}_{\text{Knowledge use (n)}}$  be the number of knowledge uses on the  $n$ th day, and the value period of the knowledge point is  $t$  (days), then the unit price of knowledge usage on the  $n$ th day is:

$$f_{\text{Knowledge unit price}}(K, n) = \frac{\sum (f_{\text{Knowledge value expectation}}(K, x) \cdot f_{\text{Knowledge use expectation}}(K, x)) / \text{Count}_{\text{Knowledge use (n-1)}}}{\sum (f_{\text{Knowledge value expectation}}(K, x_i) \cdot f_{\text{Knowledge use expectation}}(x_i))} \quad (4)$$

$$x \in \left( \frac{n-1}{10t}, \frac{n}{10t} \right], x_i \in (0, 10]$$

Each time the knowledge is used, the K-Points are weighted according to the unit price of knowledge usage.

### 3.2.2 OpenKG Token: Honor Point Measurement

OpenKG chain designed the OpenKG-Token contract to honor the knowledge contributors (publishers, reviewers, and modifiers). The OpenKG-Token is dynamically calculated and distributed to the knowledge contributors when the knowledge is used. The more knowledge is used, the more points are rewarded for its contributor. In the initial situation, the value will be equally distributed to the knowledge contributors.

$$f_{\text{Single honor value}} = \frac{f_{\text{Knowledge unit price}}}{\text{Count}_{\text{contributor}}} \quad (5)$$

The total OpenKG Token satisfies the following relationship:

$$f_{\text{Total honor value}} = g(k) \sum f_{\text{Knowledge value}} \quad (6)$$

In the initial situation,  $g(k) = 1$ .

### 3.3 Decentralized Identity Management in OpenKG Chain

The key of OpenKG chain is a trusted infrastructure. OpenKG chain adopts the VBFT consensus algorithm. Based on the traditional BFT (Byzantine Fault Tolerance) algorithm, the “VRF (Verifiable Random Function)” is introduced, which improves the anti-attack ability of the consensus algorithm and at the same time increases the consensus speed. Ontology Network uses the WasmJIT technology as the smart contract execution environment. Meanwhile it provides Layer 2 technologies to balance on-chain business performance and blockchain network expansion solutions (Figure 6).

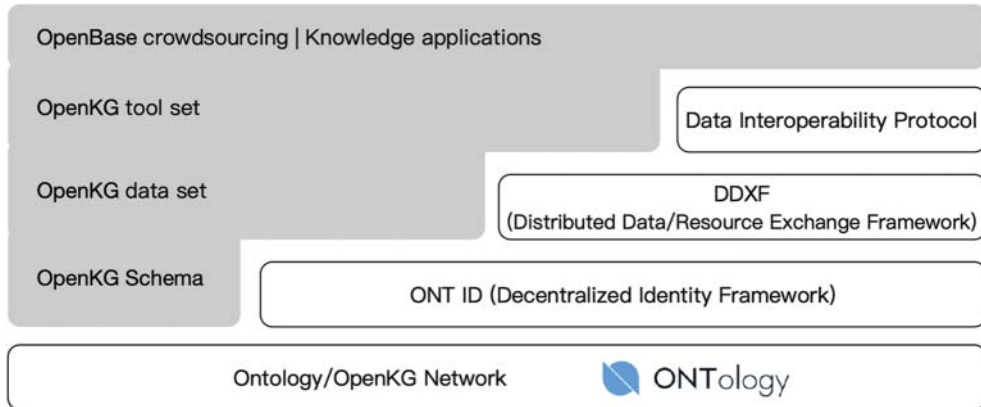


Figure 6. OpenKG chain layered architecture.

At the business application level, OpenKG chain proposes the decentralized identity identification protocol (ONT ID) for identity management in the whole lifecycle of OpenKG chain including K-Points calculation, resource management, and contributor’s identification. The distributed data exchange framework (DDXF) manages and tracks the whole process of knowledge construction, dissemination, and consumption with cross-system interoperability protocols. ONT ID can issue verifiable credentials for identifying entities, verifying credentials, supporting multi-dimensional authentication, and accessing different trusted sources. Distributed identity identification and multi-dimensional verifiable credentials provide a credible account system and risk control model for different use scenarios of knowledge.

### 3.4 Data Right Management in OpenKG Chain

The construction and use of OpenKG chain’s data involve multiple rights such as knowledge ownership, sorting, processing, viewing, and downloading. A salient challenge here is to support the right management at a different level of granularities of knowledge such as data sets, entities, and triplets. The OpenKG chain uses a distributed identity and token scheme to provide fine-grained authority management for multiple types of knowledge resources.

Firstly, OpenKG chain's data including even a single triple hold ONT ID, and the identification of data in different systems is unique. Further, for different knowledge usage scenarios, the knowledge owners and contributors can actively create knowledge authority tokens that are performed completely on the chain and guarantee safety and reliability in the whole process of token usage. Meanwhile, all OpenKG chain users also hold ONT ID, which can identify the same user in different knowledge usage scenarios of different systems and can trace back to knowledge contributors across systems to ensure the traceability of all operations. As shown in Figure 7, the specific implementation details are summarized as below:

- Data and user entities have ONT ID.
- For different scenarios, the addition, deletion, modification, and checking operations upon knowledge are managed through off-chain tokens.
- Each off-chain authority token corresponds to an on-chain data token, namely: OpenKG data-token.
- We use the property relationship between data-token and ONT ID on the chain to confirm cross-system token rights.
- Operational authentication is performed through the binding relationship between on-chain data-token and off-chain system tokens.

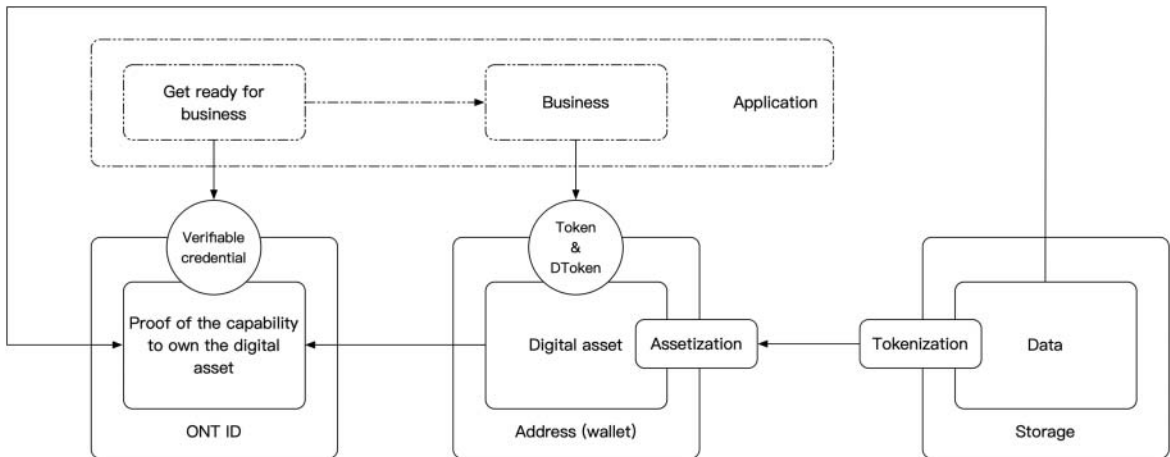


Figure 7. Data right management model of OpenKG chain.

### 3.5 Trust Management in OpenKG Chain

Why do contributors and users trust the knowledge published in OpenKG chain? For trust management, OpenKG chain provides credibility metrics for published knowledge from three levels (Figure 8):

- Infrastructure Level. The underlying network scale and node distribution of the OpenKG blockchain provides the basic endorsement of credibility of published knowledge. It will be more difficult to cheat on a network with more decentralized nodes.

- Knowledge Management Level. As all operations on knowledge in OpenKG chain are recorded on the chain, which is tamper-proof and traceable, it can provide trust endorsement for the authenticity and consistency of the data.
- Knowledge Contributor and User Level. Since all behaviors of contributors and users are also recorded and traceable on the chain, the analysis of contributors' or users' behavior can be used as a credible endorsement. It is worth mentioning that the blockchain cannot identify malicious data, but it can provide proof of malicious behavior outside the system and is permanently valid, which in turn affects the behaviors of contributors or users.

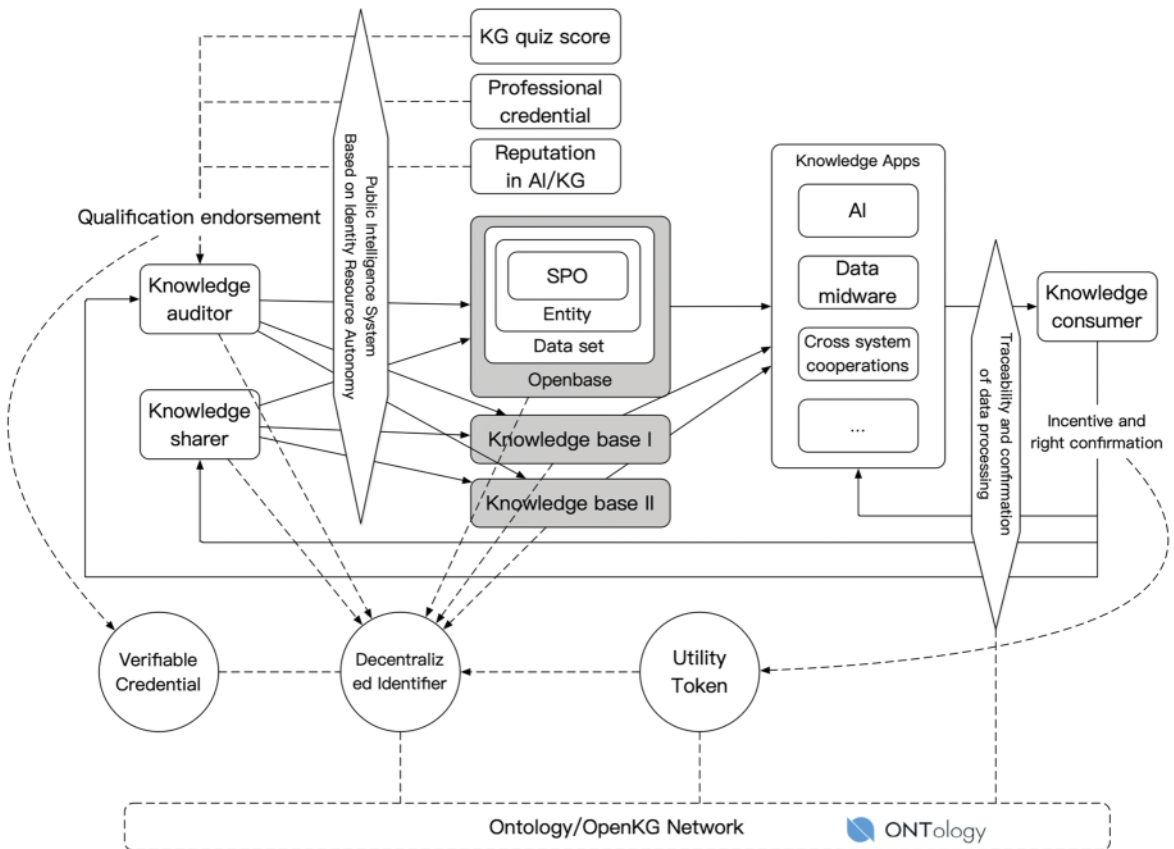


Figure 8. OpenKG chain model architecture.

## 4. IMPLEMENTATION AND EVALUATION

### 4.1 General Implementation

As introduced in the previous section, the initial nodes of the OpenKG blockchain network are tentatively set to seven, which are delivered to different institutes for operation. They form a multi-centered trusted network infrastructure upon which knowledge contributors can share data and normal users can retrieve knowledge in a safer and more accountable mode. Due to performance consideration, only operations on knowledge are recorded on chain. To synchronize on-chain operation records and off-chain knowledge, OpenKG chain implements a tokenized contract to solve the problem of data entity identification of off-chain knowledge. The whole process of using knowledge tokens on the chain is recorded to ensure the integrity of the operations while ensuring traceability. Additionally, OpenKG chain supports knowledge contributors to independently manage their knowledge data and also enables multi-party knowledge collaboration under the premise of knowledge privacy protection. In summary, the implementation of OpenKG blockchain enables the following new functionalities:

- Knowledge Index and Resource Synchronization. The on-chain records provide an index to off-chain knowledge which is stored distributively and synchronized through the blockchain.
- Safe Knowledge Consumption. Any types of knowledge consumption including browsing, downloading, and learning are recorded on the chain, ensuring safer usage and knowledge exchange.
- Accountable Knowledge Processing. Any types of operation on the knowledge including addition, audit, modification, and abolition are recorded on the chain, ensuring more accountable knowledge management.
- Knowledge Traceability. Since any operations upon knowledge are recorded on the chain, we could trace the changelog of even a single triple based on the history of the alliance chain.

### 4.2 OpenKG.CN on Chain

#### 4.2.1 Introduction to OpenKG.CN

OpenKG.CN is the main website that provides a unified sharing platform for different types of open resources. Currently it supports the sharing of open KG data and open tools. Users can freely contribute and download various types of resources on this platform. The OpenKG.CN platform currently supports three blockchain operations: user registration, resource registration and resource download. As seen in Figure 9, we build a visualization website for OpenKG community users to check their OpenKG tokens. We also make it possible to see the value of resources in Figure 10, which will be updated in real-time.

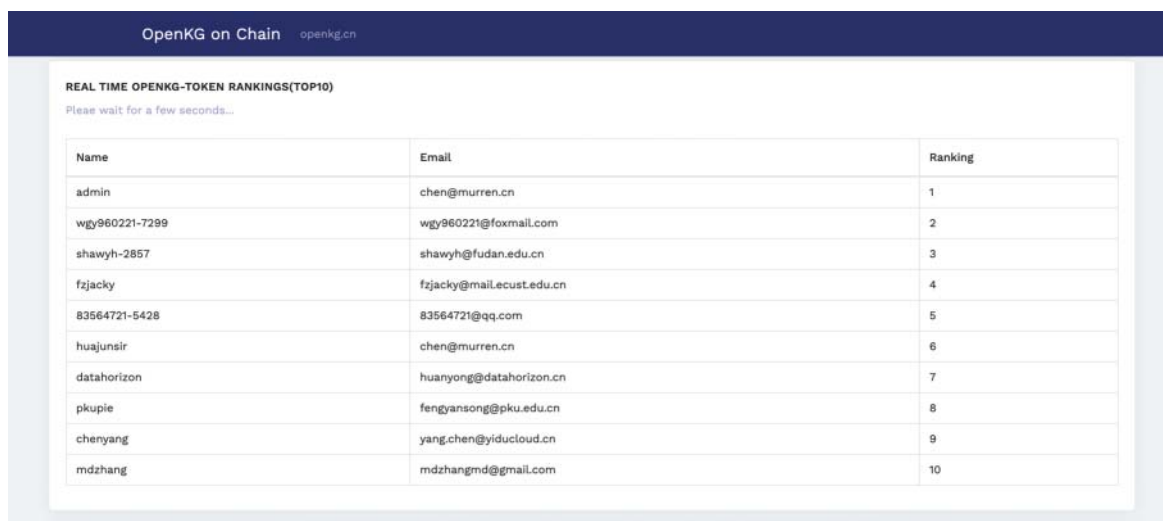
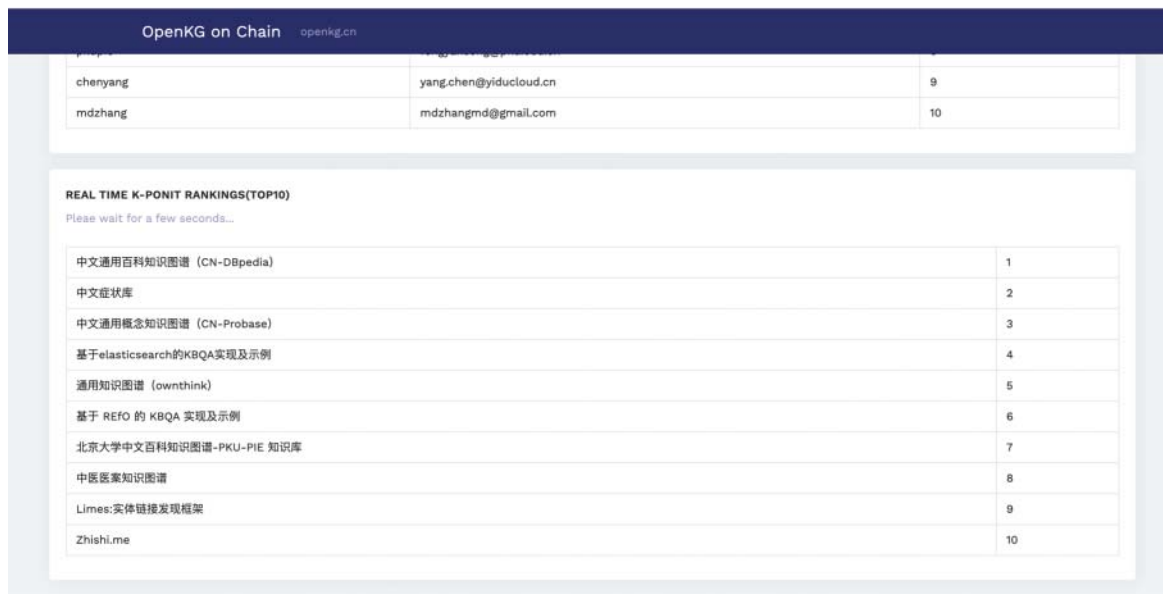


Figure 9. OpenKG-token rankings at OpenKG.CN.



© 2020 by OpenKG

Figure 10. K-point rankings at OpenKG.CN.



#### 4.2.2 Resource Registration on Chain

- User registration on the chain. When a user registers at OpenKG.CN, the system will automatically complete the registration of user information on the blockchain server and generate an on-chain account, i.e., the ONT ID as a surrogate of the user on the chain.
- Resource registration on the chain. Users can upload resources to the platform after they have registered themselves at OpenKG.CN. For each resource, the system will automatically generate a resource ID, i.e., the ONT ID for data or tools, and register the resource on the blockchain server. Please note that there is no OpenKG Token generated at this time since that value can only be generated when the knowledge is consumed.

#### 4.2.3 Resource Value Lighting-up

Since OpenKG.CN only provides coarse-grained knowledge sharing at the level of data sector tools, the main type of resource consumption and resource value lighting-up is implemented through downloading. When the resource is downloaded and used by other users, the system will generate the corresponding OpenKG Token according to the resource ID and assign it to the account of the resource contributors.

### 4.3 OpenBase on Chain

#### 4.3.1 Introduction to OpenBase

OpenBase is a crowdsourcing platform that enables fine-grained triple-level knowledge sharing within OpenKG community. The whole process includes triple addition, error checking, and knowledge graph reviewing. OpenBase takes into account the construction cost and speed of knowledge graphs at the same time. To solve the problems of fine-grained crowdsourcing construction and error checking and completion of knowledge graphs, OpenBase can take into account the construction cost and speed of knowledge graphs at the same time. It is constructed by machines and reviewed and modified by people. Targeted the existing knowledge graph, OpenBase builds a unified crowdsourcing platform for crowdsourcers to implement tasks like error checking and review of the knowledge graph.

#### 4.3.2 Fine-grained Knowledge on Chain

Traditional knowledge graph crowdsourcing platforms cannot completely solve the problem of mutual trust among users. Inspired by the idea of blockchain, we build OpenBase upon a trusted chain network to enable trusted knowledge sharing at a fine-grained triple level. Figure 11 illustrates the whole procedure of crowdsourcing triples on chain. User operations such as adding a new triple, reviewing knowledge contributed by other users, searching or querying knowledge, or downloading the entire data set will all generate related OpenKG Tokens. And all these types of user operations on data will be recorded on the blockchain for future enquiring. As to user management, when a user registers himself at OpenBase, it will be associated with an ONT ID which is decentralized managed on the underlying blockchain. Any operations issued by the user will be also associated with corresponding data and recorded on the chain.

For reward management, as per current setting, data access will not reward visitors with OpenKG Tokens, but only reward those who contribute to the initial data. The OpenKG Token will also be generated when data are reviewed or checked by reviewers. The OpenKG Token will be copied and distributed among multiple copies, which are equally divided among multiple reviewers and the original contributor. In these cases, the owner of the data is still the original user who uploaded the data. However, if an edit operation is issued, i.e., when a user modifies and edits the data, the user and the original contributor will become the owner of the data. Accessing the data (search, QA, etc.) will be regarded as a lighting-up operation, which will generate honor points to the contributors of the data set. When editing data, the editor will share ownership of the data with the original contributor. Downloading the data set will also generate OpenKG Tokens, which are divided among data contributors. The operation of adding entities and attributes will be regarded as the registration process of new data, and the operator will become the owner of the new data.

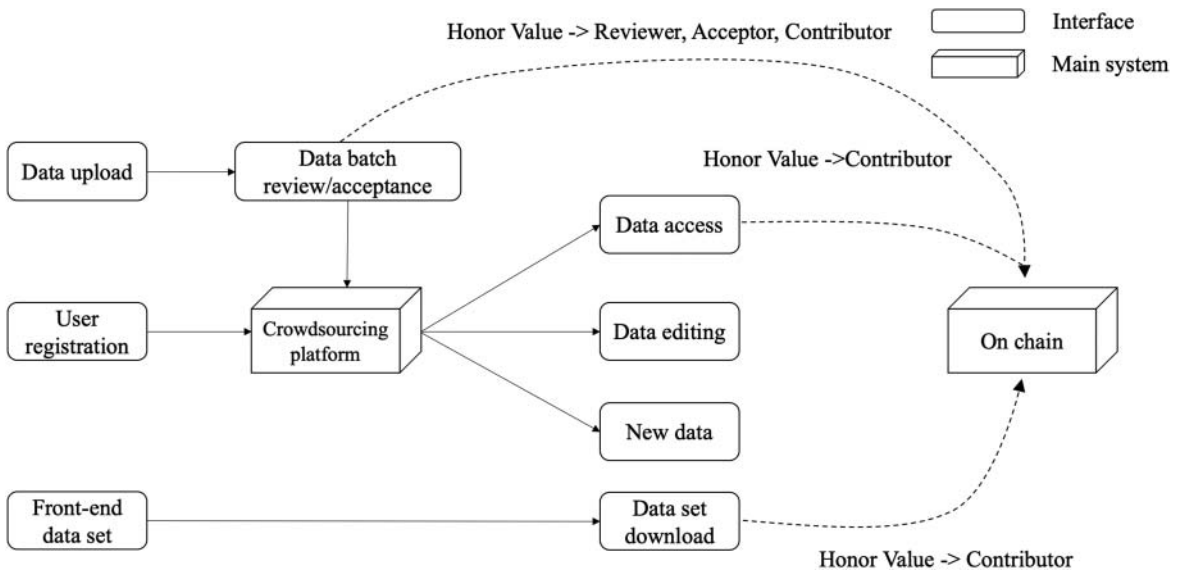


Figure 11. Illustration of OpenBase blockchain architecture.

#### 4.3.3 Lighting-up Fine-grained Knowledge

For OpenBase, there are several ways of triggering the lighting-up of the knowledge value.

- Data search and QA. When users search and query the data, the corresponding knowledge will be lighting-up and a certain amount of OpenKG tokens will be generated.
- Data downloading. When users download a data set, a certain amount of OpenKG tokens will also be generated to reward data contributors.
- Data review and checking. When users review and check data, all relevant reviewers or contributors will be rewarded with a certain amount of honor.

## 5. CONCLUSION AND FUTURE WORK

Knowledge is a valuable resource, and linking knowledge can further increase the value of knowledge. The production, transformation, exchanging, and consumption of knowledge form the value chain of knowledge in society. The value network of knowledge graph includes not only the contributors of knowledge but also the users of knowledge. The whole process of knowledge construction and consumption will gradually enrich the knowledge network and increase the knowledge value essentially. This process puts forward new requirements for rewarding knowledge at a triple granularity, self-sovereign knowledge, adversarial attack and knowledge accountability, etc. It is challenging to build a trustable value chain for the whole lifecycle of knowledge upon the open Web infrastructure.

OpenKG chain makes attempts at addressing these challenges based on the state-of-the-art blockchain technology. We hope to provide a valuable reference for the communities to help build their own enterprise-level knowledge graph crowdsourcing platform. Although the blockchain technology provides new solutions for some of the problems mentioned above, they are still not capable of solving all the problems. We still face many challenges, such as performance issues caused by fine-grained knowledge identification on the chain, decentralized storage of knowledge graphs, and trainable incentive models for knowledge crowdsourcing.

The framework of blockchain-based OpenKG chain provides a technical solution to manage the lifecycle of knowledge, as well as the process of its value discovery, standing on the perspective of knowledge itself. Furthermore, the process to form knowledge faithfully reflects the self-sovereign of user data. By bridging the physical identity with digital identity, expanding the concept of knowledge to common valuable information from Web pages, it is possible to form an Internet “twin” social network. The formation will provide effective experimental support by then. Currently, OpenKG chain implements the on-chain test of data sets, tool sets, and knowledge in the form of triplets. The methods of knowledge lighting-up are only limited to downloading and searching. In the future, we will try more diverse types of resources, including KG schemas, bots, and knowledge graph algorithms. We will also explore richer modes of knowledge lighting-up such as question answering, decentralized reasoning, and federated knowledge learning.

## AUTHOR CONTRIBUTIONS

H.J. Chen (huajunsir@zju.edu.cn), N. Hu (huning@ont.io), G.L. Qi (gqi@seu.edu.cn) and H.F. Wang (wang\_haofen@gowild.cn) designed the framework between open knowledge graphs and blockchain infrastructure, and summarized the discussion part of this paper. B. Zhen (bizhen\_zju@zju.edu.cn), F. Yang (294948563@qq.com) and J. Li (lijie@onchain.com) lead the implementation of OpenKG chain in the OpenKG community. All the authors have made meaningful and valuable contributions in revising and proofreading the resulting manuscript.

## REFERENCES

- [1] Berners-Lee, T., Hendler, J.: Publishing on the semantic Web. *Nature* 410, 1023–1024 (2001)
- [2] Wang, Z., et al.: Knowledge graph embedding by translating on hyperplanes. In: *The 28th AAAI Conference on Artificial Intelligence*, pp. 1112–1119 (2014)
- [3] Berners-Lee, T., Fischetti, M.: *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. DIANE Publishing Company, Darby (2001)
- [4] Berners-Lee, T., et al.: World-Wide Web: The information universe. *Internet Research* 20(1), 461–471 (2010)
- [5] Berners-Lee, T., Hendler, J., Lassila, O.: The semantic Web. *Scientific American* 284, 34–43 (2001)
- [6] Zhang, Y., et al.: Variational reasoning for question answering with knowledge graph. *arXiv preprint arXiv:1709.04071* (2017)
- [7] Chen, Y.-N., et al.: Matrix factorization with knowledge graph propagation for unsupervised spoken language understanding. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 483–494 (2015)
- [8] Zhang, Z., et al.: Ernie: Enhanced language representation with informative entities. *arXiv preprint arXiv:1905.07129* (2019)
- [9] Schlichtkrull, M.: Welling, modeling relational data with graph convolutional networks. In: *European Semantic Web Conference*, pp. 593–607 (2018)
- [10] Lin, H., et al.: Learning entity and relation embeddings for knowledge resolution. *Procedia Computer Science* 108, 345–354 (2017)
- [11] Wang, X., et al.: Kgat: Knowledge graph attention network for recommendation. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 950–958 (2019)
- [12] Guarino, N.: Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human Computer Studies* 43, 625–640 (1995)
- [13] Gruber, T.R.: The role of common ontology in achieving sharable, reusable knowledge bases. In: *Proceedings of KR' 1991*, pp. 601–602 (1991)
- [14] Guo, Y., Pan, Z., Heflin, J.: Lubm: A benchmark for OWL knowledge base systems. *Journal of Web Semantics* 3, 158–182 (2005)
- [15] Crane, S.: Networked knowledge representation and exchange using UML and RDF. *Journal of Digital Information* 1 (8) (2001)
- [16] Velicković, P., et al.: Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017)
- [17] Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: *Advances in Neural Information Processing Systems*, pp. 1024–1034 (2017)
- [18] Liu, H., Singh, P.: Conceptnet—A practical commonsense reasoning tool-kit. *BT Technology Journal* 22, 211–226 (2004)
- [19] Vrandečić, D., Krötzsch, M.: Wikidata: A free collaborative knowledgebase. *Communications of the ACM* 57, 78–85 (2014)
- [20] Coppens, S., et al.: Reasoning over SPARQL. In: *LDOW*, pp. 1–5 (2013)
- [21] Aasman, J.: Utilizing federated knowledge in semantic web applications. In: *2008 IEEE International Conference on Semantic Computing*, pp. 486–487 (2008)
- [22] Bao, J., Caragea, D., Honavar, V.: A tableau-based federated reasoning algorithm for modular ontologies. In: *2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)* (WI'06), pp. 404–410 (2006)

- [23] Konečný, J., et al.: Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492 (2016)
- [24] Chen, M., et al.: Fede: Embedding knowledge graphs in federated setting. arXiv preprint arXiv:2010.12882 (2020)
- [25] Swan, M.: Blockchain: Blueprint for a new economy. O'Reilly Media, Boston (2015)
- [26] Crosby, M., et al.: Blockchain technology: Beyond bitcoin. *Applied Innovation* 2, 71 (2016)
- [27] Ølnes, S., Ubacht, J., Janssen, M.: Blockchain in government: Benefits and implications of distributed ledger technology for information sharing. *Government Information Quarterly* 34(3), 355–364 (2017)
- [28] Mills, D.C., et al.: Distributed ledger technology in payments, clearing, and settlement (2016). Finance and Economics Discussion Series 2016-095. Washington: Board of Governors of the Federal Reserve System. Available at: <https://doi.org/10.17016/FEDS.2016.095>. Accessed 27 April 2021.
- [29] Davidson, S., De Filippi, P., Potts, J.: Disrupting governance: The new institutional economics of distributed ledger technology. Available at: <http://dx.doi.org/10.2139/ssrn.2811995>. Accessed 27 April 2021
- [30] Mingxiao, D., et al.: A review on consensus algorithm of blockchain. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2567–2572 (2017)
- [31] Bach, L., Mihaljevic, B., Zagar, M.: Comparative analysis of blockchain consensus algorithms. In: The 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1545–1550 (2018)
- [32] Nguyen, G.-T., Kim, K.: A survey about consensus algorithms used in blockchain. *Journal of Information Processing Systems* 14(1), 101–128 (2018)
- [33] van der Aalst, W.M., Bichler, M., Heinzl, A.: Robotic process automation. *Business & Information Systems Engineering* 60(4), 269–272 (2018)
- [34] Willcocks, L.P., Lacity, M., Craig, A.: The IT function and robotic process automation. LSE Research Online Documents on Economics 64519, London School of Economics and Political Science, LSE Library (2015)
- [35] Yang, J., Adamic, L.A., Ackerman, M.S.: Crowdsourcing and knowledge sharing: Strategic user behavior on taskcn. In: Proceedings of the 9th ACM conference on Electronic commerce, pp. 246–255 (2008)
- [36] Chu, X., et al.: Katara: A data cleaning system powered by knowledge bases and crowdsourcing. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, pp. 1247–1261 (2015)
- [37] Roy, S.B., et al.: Task assignment optimization in knowledge-intensive crowdsourcing. *The VLDB Journal* 24, 467–491 (2015)

## AUTHOR BIOGRAPHY



**Huajun Chen**, Full Professor of College of Computer Science and Technologies at Zhejiang University, serves as the Director of Joint Lab on Knowledge Engine of Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies (AZFT), and a Deputy Director of the Key Lab of Big Data Intelligence at Zhejiang Province. His research interests include knowledge graph technologies, ontologies and the semantic Web, information extraction and natural language processing, and applications in areas such as biomedicine, e-commerce and smart cities.

ORCID: 0000-0001-5496-7442



**Ning Hu**, Chief of Technology at Ontology Foundation, the high performance, public blockchain specializing in decentralized identity and data. He has almost two decades of experience in software development and high-end technology, including semantic Web, big data and artificial intelligence (AI).

ORCID: 0000-0002-9598-8554



**Guilin Qi** is a Professor at Southeast University, China, where he also serves as Director of the Institute of Cognitive Intelligence and of the Knowledge Science and Engineering Lab. His research interests include knowledge representation and reasoning, knowledge graphs, uncertainty reasoning, and the semantic Web. Prof. Qi is an editorial board member of the *Journal of Web Semantics*, and has co-edited special issues for the *Annals of Mathematics and Artificial Intelligence*, *International Journal of Approximate Reasoning* and *Journal of Applied Logic*. He has over 20 years of research experiences in knowledge engineering and has led many national and industrial projects on knowledge graphs. Prof. Qi has published more than 100 papers on knowledge engineering and knowledge graphs and holds two patents. He has given many invited talks about knowledge graphs and has given lectures on knowledge graph techniques to postgraduate students. He is one of the leading experts on knowledge graphs in China.

ORCID: 0000-0002-1957-6961





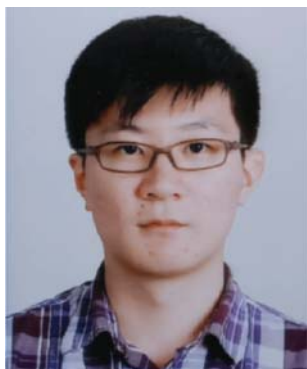
**Haofen Wang** has long served as Chief Technology Officer in a first-line artificial intelligence company and has rich experience in AI R&D management. He is one of the founders of OpenKG, the world's largest Chinese open knowledge graph alliance. He is responsible for participating in a number of provincial and ministerial AI-related projects, and has published more than 90 high-level papers in the AI field, which have been cited more than 2,000 times and the H-Index has reached 21. He built the world's first interactive virtual idol—"Amber·Xuyan"; the intelligent customer service robot he built has served more than 1 billion users. Currently, he is the Deputy Director of the Terminology Committee of China Computer Society, the Deputy Secretary General of the Language and Knowledge Computing Committee of Chinese Information Society and distinguished research fellow in College of Design Innovation, Tongji University.

ORCID: 0000-0002-0672-8081



**Zhen Bi** is a graduate student from Zhejiang University. His research interests include natural language processing and knowledge graph.

ORCID: 0000-0002-3287-5683



**Jie Li** is currently a Back-end Developer at Ontology Foundation. His work is devoted to the development of data and identity information based on the blockchain technology.





**Fan Yang** is a graduate student from Zhejiang University. His research interests include natural language processing and knowledge graph.  
ORCID: 0000-0002-4577-8196